AD-A091 553 DESMATICS INC STATE COLLEGE PA F/G 9/2
AN EXAMINATION OF THE PERFORMANCE OF TWO-STAGE GROUP SCREENING --ET
OCT 80 C A MAURO, D E SMITH N00014-79-C-0650
UNCLASSIFIED TR-113-3 NL

1 OF 1
AD
A091553

END
DATE
FILMED
12-80
DTIC

C

AN EXAMINATION OF THE PERFORMANCE OF TWO-STAGE GROUP SCREENING FOR USE IN COMPUTER SIMULATION

by

10

Carl A. Mauro

[redacted]

Dennis E. Smith

DESMATICS, INC.

12

P. O. Box 618
State College, Pa. 16801
Phone: (814) 238-9621

Applied Research in Statistics - Mathematics - Operations Research

6

AN EXAMINATION OF THE PERFORMANCE OF TWO-STAGE GROUP SCREENING FOR USE IN COMPUTER SIMULATION

by

10

Carl A. Mauro

[redacted]

Dennis E. Smith

14 TR-

9

TECHNICAL REPORT NO. 113-3

11

Oct[redacted]80

12 57

This study was supported by the Office of Naval Research under Contract No. N00014-79-C-0650 Task No. NR 274-317

15

391 156

ABSTRACT

When only a limited number of simulation runs are available, it is usually impossible to thoroughly investigate all of the factors under consideration. Often, though, it is anticipated that only a small subset of the original factors is important. Accordingly, it is desired to screen the factors in order to help identify those that do exert an appreciable effect on the simulation response. Two-stage group screening is one possible strategy that may be used. However, a basic assumption of this strategy is that the directions of possible effects are known, a priori. This report examines, in the case of zero error variance (i.e., when the simulation response is observed without random error), the performance of two-stage group screening when the assumption of known directions is relaxed.

Key Words:

Factor Screening

Two-Stage Group Screening

Computer Simulation

TABLE OF CONTENTS

I. INTRODUCTION

Many computer simulation studies involve a large number of factors (i.e., independent variables), but are limited in the number of simulation runs available. In these cases, a major problem that a simulation user faces is to determine the effects of the factors on the simulation response (i.e., dependent variable).

When only a limited number of simulation runs are available, it is usually impossible to thoroughly investigate all the factors. Often, though, it is anticipated that only a small subset of the many original factors is important. Accordingly, it is desired to screen the factors in order to help identify the relatively few that do exert an appreciable effect on the simulation response. Such experiments are called factor screening experiments, and the problem of experimentation to determine the size and the composition of the subset of important factors is called the factor screening problem.

Although this report places the factor screening problem in a computer simulation context, this type of problem can arise in practically any field of scientific research in which conventional designs are prohibited due to the number of experimental observations they require. Therefore, the results of this report could, if desired, be generalized to embrace a wider experimental framework or application.

For the purpose of discussion, define as active any factor which produces a nonzero change in the mean simulation response. Hence, every factor can be classified as an active or an inactive factor. In general, therefore, the basic objective of any factor screening strategy is to efficiently and effectively classify the factors under investigation. Of course, these notions

will need to be made precise in order to objectively evaluate the performance of a screening strategy.

Unfortunately, to date, no specific methodology has been developed to adequately resolve the factor screening problem, although a number of general suggestions have been made [e.g., Hunter and Naylor (1970), Jacoby and Harrison (1962), Kleijnen (1945), Smith (1973), Montgomery (1979), Smith and Mauro (1980)]. Despite these suggestions, there remains a lack of analytical and empirical results with which to systematically analyze the factor screening problem.

II. BACKGROUND

Smith and Mauro (1980) suggested two screening strategies for further research and also suggested a performance measure to evaluate them (or any other strategy). The two screening strategies identified as warranting more detailed investigation were the two-stage group screening method and a random balance/Plackett-Burman design two-stage combination strategy. Both of these strategies are intended for use in situations in which the available number of computer runs is less than the number of factors to be screened.

A basic assumption, however, in the group screening strategy is that the directions of possible effects are known, a priori. Historically [Watson (1961), Kleijnen (1975)], the effect of this assumption on performance has been regarded with little concern. The primary objective of this report is to explicitly examine, in the case of zero error variance (i.e., when the simulation response is observed without random error), the performance of two-stage group screening when the assumption of known directions is relaxed. The results of this analysis should help determine whether group screening can be effectively used even when all the directions of potential effects are not known (or are incorrectly specified).

Initially, the following factor screening model will be assumed to underlie the simulation responses. In section IV two further conditions will be imposed on this model. The factor screening model is

$$y_i = \beta_0 + \sum_{j=1}^{K} \beta_j x_{ij} + \epsilon_i, \qquad (2.1)$$

where

(1) y_i is the response of the i^{th} simulation run,

(2) K is the total number of factors to be screened, each of which is at two levels,

(3) x_{ij} is +1 or -1 depending on the level of the j^{th} factor for the i^{th} simulation run,

(4) β_j is the (linear) effect of the j^{th} factor,

and (5) the error terms, ε_i, are independent with common distribution $N(0, \sigma^2)$.

III. GROUP SCREENING

In the first stage of the two-stage group screening procedure, originally proposed by Watson (1961), group-factors are formed by subdividing the original factors into groups of (roughly) equal size. The group factors are considered as single factors and are examined in a Plackett-Burman design.[1] Group-factor levels are defined by equating the levels of component factors to the group-factor level. In the second stage of the procedure, the individual factors which comprise the significant group-factors are also examined in a Plackett-Burman (1946) design.

Since the grouping procedure induces complete confounding among the factors within a group, the second stage can be regarded as a follow-up experiment that separates the effects of factors which comprise the significant group-factors. A brief description of the two-stage group screening procedure is outlined in the following paragraphs.

A. STAGE ONE

The initial step is to randomly partition the K factors into groups of size g. To allow for general g, denote this particular variation as the GS_g strategy. If K is not a multiple of g, then the group sizes should be taken as "evenly" as possible. For example, if K = 46 and g = 3, the partition would consist of 14 groups of size 3 and 1 group of size 4, for a

[1] Plackett-Burman designs are two-level, orthogonal fractional factorial designs for studying up to (4m - 1) factors in 4m runs. [See Plackett and Burman (1946).]

total of 15 groups. (See section VI for practical considerations in factor partitioning.) Let G denote the number of groups into which the K factors are partitioned.

Once the factors are partitioned into G groups, the group-factors are systematically tested in a Plackett-Burman design employing N runs. If B(x) denotes the smallest integer larger than x that is a multiple of four, then N = B(G). For instance, in the above example, N = B(15) = 16.

The +1 level of a group-factor is defined as the level in which all factors within that group are maintained at their +1 level. The -1 level of a group-factor is analogously defined. However, when specifying the original factor levels, the level which will produce the larger response (if, in fact, the factor is active) should be defined as the +1 level in order to prevent the possibility of group-factor effect cancellation. Thus, all factors which are active will have effects with the same direction. As mentioned previously, it has been assumed in the development of two-stage group screening that this can be done. This report examines the impact of relaxing this assumption.

B. STAGE TWO

Only the individual factors which comprise the group-factors determined to have significant effects in stage one are carried over into this stage. The factors reaching this stage are tested individually in a Plackett-Burman design. If s equals the number of factors carried over, then the required number of second stage runs, M, equals B(s).

Any factor determined to have a significant effect in this stage is classified as active. Any factor not having a significant effect or any factor not even carried over from stage one is classified as an inactive factor. The total number of runs, R, that the two stages require will equal $N + M = B(G) + B(s)$.

IV. ADDITIONAL ASSUMPTIONS AND NOTATION

In his development of group screening, Watson (1961) imposed two additional assumptions on the first-order model (2.1) defined previously. In addition to conditions (1)-(5), these are:

(6) all factors have, independently, the same prior probability, p, of being active,

and (7) active factors have the same linear effect, $\Delta > 0$, i.e.,

$$\beta_j = \begin{cases} \Delta \text{ , if } j^{th} \text{ factor active} \\ 0 \text{ , if } j^{th} \text{ factor inactive.} \end{cases}$$

A prerequisite for assumption (7) is that the directions of all potential effects are known. In this case, the upper level (+1) of each factor can be defined as the level which will lead to a larger response if the factor is active. Hence, the associated linear effect, β_j, will be nonnegative. More importantly, assumption (7) serves to insure against possible cancellation of effects that may occur within a group-factor.

Assumption (6) effectively states that the number of active factors is a binomial random variable with parameters K and p. The resulting model is, therefore, a type of random effects model. However, in most practical factor screening situations, a fixed effects model would appear to be the more appropriate model. In other words it would be assumed that some fixed (but unknown) number, k, of the K factors are active.

It is the purpose of this paper to examine, in the case of zero error

variance ($\sigma^2 = 0$), the performance of two-stage group screening when the assumption of known directions is relaxed and when the random effects model of assumption (6) is replaced with a fixed effects model. Thus, in place of conditions (5), (6), and (7), these modified conditions will be assumed:

(5') the observations are observed without random error (i.e., $\sigma = 0$),

(6') there are a total of K factors to be screened, $k \geq 1$ (unknown) of which are active and $(K - k)$ of which are inactive (let $p^* = k/K$),

and (7') all active factors have the same absolute linear effect, $\Delta > 0$, i.e.,

$$|\beta_j| = \begin{cases} \Delta \text{ , if } j^{th} \text{ factor active} \\ 0 \text{ , if } j^{th} \text{ factor inactive.} \end{cases}$$

Conceptually, (6') and (7') compare with assumptions (6) and (7) in the following ways:

(a) Assumption (6) states that whether or not a factor will be active is a random occurrence in which all factors have, independently, the same prior probability of being active. Thus, $k_w \sim b(K, p)$ where k_w denotes the number of active factors. Because of the symmetry in factor effect magnitude, for a fixed K it is the number of active factors and not which particular factors are active that is the essential element in formalizing the model.

In passing from (6) to (6') or from (6') to (6), assume

that $p = p^*$. Hence, $k_w \sim b(K, k/K)$, from which it follows that $E(k_w) = k$. Therefore, on the average, the number of active factors corresponding to assumptions (6) and (6') is the same.

(b) Assumption (7) requires that the directions of possible effects are known. Therefore, factor levels can be judiciously specified so that all active factors will have positive effects. Of course, it is further assumed that these effects are of the same magnitude. Since condition (7') requires only equal absolute effects, (7') relaxes the assumption of known directions. Thus condition (7) is actually a special case of (7').

For purposes of analysis, it can be assumed without loss of generality that the first k indexed factors are active and the remaining (K - k) factors are inactive. As a result of (5'), (6'), and (7'), model (2.1) can be updated and rewritten as

$$y_i = \beta_0 + \sum_{j=1}^{k} \beta_j x_{ij} \tag{4.1}$$

where x_{ij} is ± 1 depending on the level specification of factor j (j = 1, 2, ..., K) and $|\beta_j| = \Delta$ when $1 \leq j \leq k$. It is implicit that $\beta_j = 0$ when $j > k$.

If the factor level (of active factors) producing the larger response is defined as the +1 level, then $\beta_j = \Delta > 0$ for all $1 \leq j \leq k$. However, if this arrangement is not followed, then a mixture of $\pm\Delta$'s will result within $\beta_1, \beta_2, \ldots, \beta_k$. To represent this possibility, let $\underline{\beta} = (\beta_1, \beta_2, \ldots, \beta_k)'$ and let $\underline{\beta}(i)$ denote the case where $\beta_1, \beta_2, \ldots, \beta_i$ each equal $-\Delta$ and $\beta_{i+1}, \beta_{i+2}, \ldots, \beta_k$ each equal $+\Delta$. Define $\underline{\beta}(0) = (\Delta, \Delta, \ldots, \Delta)'$. Note that $\underline{\beta}(0)$ represents the case assumed in condition (7).

V. ASSESSING PERFORMANCE

Because of the first-order structure of model (2.1) and because of the resolution properties of Plackett-Burman designs, under the assumption of zero error variance, any factor reaching stage two of the GS_g strategy will be correctly classified. Consequently, in the case of zero error variance, no inactive factor will be misclassified in the GS_g strategy regardless of g or $\underline{\beta}$.

However, this is not necessarily true of the active factors. For instance, in the cases $\underline{\beta}(0)$ and $\underline{\beta}(k)$, all active factors will be correctly classified since cancellation of effects cannot occur, but in the other $\underline{\beta}(i)$ $(0 < i < k)$ arrangements, some active factors may fail to reach the second stage due to cancellation of group-factor effects. As an example, suppose four active factors, a, b, c, and d, fall into a particular group along with some inactive factors. Suppose $\beta_a = \beta_b = \Delta$ and $\beta_c = \beta_d = -\Delta$. In this case, the overall group-factor effect will total zero. Thus factors a, b, c, and d will not reach stage two and will consequently be classified as inactive factors.

Of interest, therefore, is the impact of the possibility of active factor cancellation on the performance of the GS_g strategy. In order to quantitatively assess the extent of this impact, some measure of performance needs to be defined. One possible approach to devising a performance measure was outlined by Smith and Mauro (1980). In this development, evaluating performance is essentially reduced to balancing two separate and conflicting requirements.

The first of these two requirements concerns the severity of factor

classification errors that may occur in factor screening. The following loss function was suggested for measuring this aspect of performance:

$$L(\underline{\beta}) = \sum_{j=1}^{K} w_j \delta_j / \sum_{j=1}^{K} w_j \ , \qquad (5.1)$$

where

$$\delta_j = \begin{cases} 0 \ , \text{ if the } j^{th} \text{ factor is correctly identified,} \\ 1 \ , \text{ if the } j^{th} \text{ factor is incorrectly identified,} \end{cases}$$

and w_j denotes the loss incurred if the j^{th} factor is misclassified.

For model (4.1), it is reasonable to let

$$w_j = \begin{cases} \tfrac{1}{2}k & 1 \leq j \leq k \\ \tfrac{1}{2}(K - k) & j > k \ , \end{cases}$$

since this apportions one-half of the overall maximum loss to the active factors and the other half to the inactive factors. Thus, (5.1) reduces to

$$L = [(K - k)(k - A) + k(K - k - I)]/2k(K - k) \qquad (5.2)$$

where A denotes the number of active factors correctly identified and I denotes the number of inactive factors correctly identified.

Since $L (0 \leq L \leq 1)$ is random for most screening strategies, expected loss, E(L), can be used as one basis for performance. The quantity [1 - E(L)] is, therefore, a measure of the efficiency of a screening strategy for classifying factors. Classification efficiency closer to one (or expected loss closer to zero) would indicate better performance on the average.

The other aspect of performance discussed by Smith and Mauro (1980) takes into account the total number of runs, R, that a strategy requires. Equation (5.2) alone fails to do this. Regarding this aspect of performance,

$$Q = E(R)/B(K) \tag{5.3}$$

can be used as a measure of the relative testing cost of a strategy (relative to the runs required for a Plackett-Burman design).

Although a smaller value of Q would indicate better performance on the average, it is imperative that expected loss and relative testing cost be considered jointly in assessing the overall performance of a screening strategy. As an illustration, consider the strategy employing R = 0 runs in which all factors are declared active (or inactive). In this trivial strategy, Q = 0, but classification efficiency is only 0.5!

For any strategy having acceptable expected loss and relative testing cost, (5.2) and (5.3) might then be combined into some overall measure, say H, to further evaluate and compare performance. For instance, let

$$H = [1 - E(L)]/Q. \tag{5.4}$$

A larger H would indicate better performance on the average. Other possibilities for combining (5.2) and (5.3) into an overall performance measure are discussed in Smith and Mauro (1980).

Since I = K - k in the GS_g strategy, performance (5.4) of the GS_g strategy given $\underline{\beta}(i)$ is defined and denoted

$$H_g(\underline{\beta}(i)) = \left[\frac{B(K)}{E(R|\underline{\beta}(i))}\right]\left[\frac{k + E(A|\underline{\beta}(i))}{2k}\right]. \tag{5.5}$$

Although the exact probability distributions of A and R are intractable for general $\underline{\beta}(i)$, their expected values do admit to an analytical solution (see Appendix A). However, it is necessary to assume that K is a multiple of g. In cases where K is not a multiple of g, but the group sizes are specified as "evenly" as possible, the expressions should provide reasonable

approximations to the actual moments.

Since $H_g(\underline{\beta}(i)) = H_g(\underline{\beta}(k - i))$, performance need only be evaluated for $0 \leq i \leq [k/2]$. Furthermore, in the case $i = 0$, $E(A|\underline{\beta}(0)) = k$, so that

$$H_g(\underline{\beta}(0)) = B(K)/E(R|\underline{\beta}(0)). \qquad (5.6)$$

Therefore, the group size that maximizes $H_g(\underline{\beta}(0))$ is equivalent to the group size that minimizes $E(R|\underline{\beta}(0))$.

In the case of zero error variance, minimizing the expected total number of runs is precisely Watson's condition for optimum group size under assumptions (6) and (7). Figure 1 contains Watson's (approximate) optimum group size as a function of p (the probability a factor is active) in the case of zero error variance.

Because the expected total number of runs given $\underline{\beta}(0)$ will differ under assumptions (6) and (6') for a given group size, let $E_w(R|\underline{\beta}(0))$ denote the expected total number of runs given $\underline{\beta}(0)$ under assumption (6). It is easy to show that for a partition of the K factors into G groups of sizes $g_1, g_2, \ldots, g_G$, $E_w(R|\underline{\beta}(0))$ is roughly[1]

$$B(G) + K - \sum_{i=1}^{G} g_i(1 - p)^{g_i} + 2.5. \qquad (5.7)$$

If $g = g_i$ for all i, then (5.7) reduces to $B(G) + K[1 - (1 - p)^g] + 2.5$.

It will be seen later that (5.7) with $p = p^*$ agrees fairly closely with $E(R|\underline{\beta}(0))$, although it is typically slightly smaller. Hence, in practice, $E_w(R|\underline{\beta}(0))$ can be used as a first approximation to $E(R|\underline{\beta}(0))$.

---

[1] This differs somewhat from Watson's (1961) approximation because his approximation assumes $N = G + 1$ and $M = s$.

Prior Probability p	Optimum Group Size
.01	11
.02	8
.03	6
.04	6
.05	5
.06	5
.07	5
.08	4
.09	4
.10	4
.12	4
.13 - .30	3

Figure 1: Watson's approximate optimum group size as a function of p in the case of zero error variance [under assumptions (6) and (7)].

VI. PERFORMANCE EVALUATION OF THE GS_g STRATEGY

The performance (5.5) of the $GS_g(3 \leq g \leq 8)$ strategy given $\underline{\beta}(i)$ $(0 \leq i \leq [k/2])$ was evaluated for the twelve specific cases of (K, k) listed in Figure 2. For each value of K (60, 120, and 240) under investigation, four values of k were examined. These were determined by letting $k = p^*K$, where $p^* = 2/60, 3/60, 5/60$, and $8/60$.

The number of groups into which the K factors were partitioned for the various group sizes are displayed in Figure 3. When K is a multiple of g, it is assumed that the factors are partitioned into $K/g = G$ groups of size g. In practice, this may or may not be a desirable convention. For instance, when $K = 60$ and $g = 3$, 24 first stage runs are required to test the 20 group-factors. This would seem to be a waste of four runs. An alternative plan would be to employ 16 groups of size three and three groups of size four, for a total of 19 group-factors. This partition would require only 20 first stage runs and would have a comparable expected loss.

Since the formulas for E(R) and E(A) in Appendix A assume that K is a multiple of g, such practical considerations in the number of first stage runs are not incorporated into the computations. However, the potential reduction in E(R) for GS_g strategies where these considerations are relevant can be readily approximated by subtracting 4 from the expected total number of runs calculated when K is assumed a multiple of g. In terms of Q, this modification will result in little reduction in relative testing cost unless K is small.

For each of the twelve cases, E(A), E(R), E(L), Q and H were computed

K (No. of factors)	p* (Proportion of active factors)			
	$2/60 = .03\overline{3}$	$3/60 = .050$	$5/60 = .08\overline{3}$	$8/60 = .13\overline{3}$
60	2	3	5	8
120	4	6	10	16
240	8	12	20	32

Figure 2: Twelve cases examined. (Entry is k = p*K, the number of active factors.)

K (No. of factors)	g (Group size) 3	4	5	6	7	8
60	20	15	12	10	9*	7*
120	40	30	24	20	17*	15
240	80	60	48	40	34*	30

*For these cases K is not a multiple of g. However, $E(A)$ and $E(R)$ are calculated as if $K = Gg$. For instance, when $K = 120$ and $g = 7$, $E(A)$ and $E(R)$ are calculated as if there were 17 groups of size 7, for a total of 119 factors.

Figure 3: The number of group-factors, G.

for every GS_g ($3 \leq g \leq 8$) strategy and for each $\underline{\beta}(i)(0 \leq i \leq [k/2])$ case. The complete output appears in Appendix B. The results of this analysis pertaining to loss, relative testing cost, and performance are summarized below.

A. LOSS

In general, expected loss is smaller for a smaller group size. Furthermore, for any group size, $E(L|\underline{\beta}(i)) = 0$ at $i = 0$ and increases to a maximum value at $i = [k/2]$.

To better understand what loss represents, it is instructive to consider the quantity $[1 - 2E(L)]$. Recall that in the GS_g strategies, $E(L) = (k - E(A))/2k$, hence $1 - 2E(L) = E(A)/k$. Therefore, in this particular case (model 4.1), $1 - 2E(L)$ represents the proportion of the active factors that are correctly identified. Hence, to say that expected loss increases is to say that the expected proportion of active factors that are correctly identified decreases and at twice the rate. Note that if $E(A)/k$ is desired to be greater than γ, then $E(L)$ must be smaller than $(1 - \gamma)/2$.

As a further aid, define $E_c(L) = E(L|\underline{\beta}(ck))$, $0 \leq c \leq 1$. The argument c can be interpreted as the proportion of the active factors having an effect of $-\Delta$. For the moment, suppose that $E_c(L)$ is a continuous function in c. (Use linear interpolation to define $E_c(L)$ when ck is not an integer.)

For a fixed group size, note that $E_0(L) = 0$, $0 \leq E_c(L) \leq 1$, $E_c(L)$ is symmetric about .5 and is an increasing function over the interval $0 \leq c \leq .5$. Furthermore, for a fixed $c > 0$ and g, $E_c(L)$ increases as $p^* = k/K$ increases. That is, given a fixed proportion of active factors equal to $-\Delta$, expected

loss will increase as the proportion of active factors increases. For fixed p* and g, $E_c(L)$ is essentially independent of K, for $60 \leq K \leq 240$, at least when g is small. (See Appendix C for the case p* = .05.)

Figure 4 presents the $E_c(L)$ curves corresponding to p* = .05 and K = 240. Curves could be constructed for the other p* values by using the data given in Appendix B. These curves are useful in determining what combinations of group size and $\underline{\beta}(i)$, for a given p*, will meet (i.e., not exceed) a prespecified maximum tolerated loss.

B. RELATIVE TESTING COST

As expected, due to the increasing chance of group-factor effect cancellation, relative testing cost will decrease as the proportion of active factors having an effect of $-\Delta$ increases from 0% to 50%. In terms of group size, for a fixed $\underline{\beta}(i)$, there is a group size that minimizes Q (relative testing cost) such that using a larger or smaller group size will only serve to increase the testing cost. For an example, see Figure 5.

It is interesting, however, to observe (at least for the range of parameter values considered here) that the group size that minimizes Q at $\underline{\beta}(0)$ also minimizes Q at the other $\underline{\beta}(i)$, $i > 0$, arrangements. This is apparent upon inspection.

It was also found that equation (5.7) with p = p* provides a reasonable approximation to $E(R|\underline{\beta}(0))$. See Figure 6 for four specific cases. Furthermore, as already noted, Q decreases as i increases from 0 to $[k/2]$. As a first approximation to $E(R|\underline{\beta}([k/2]))$, an empirically derived formula can be used. This formula, obtained by using least-squares applied to the data in

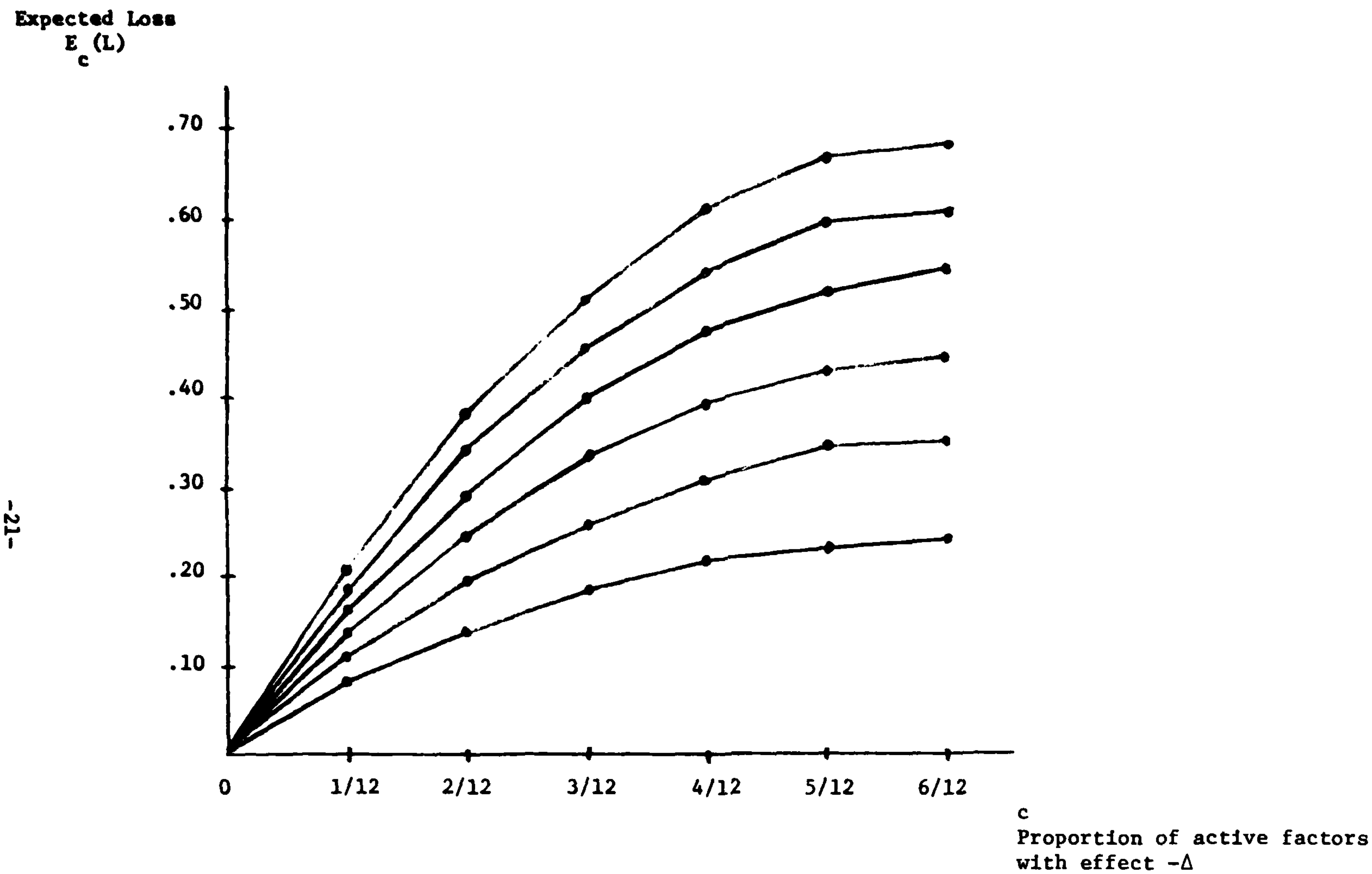

Figure 4: $E_c(L)$ curves corresponding to $p^* = .05$ and $K = 240$.

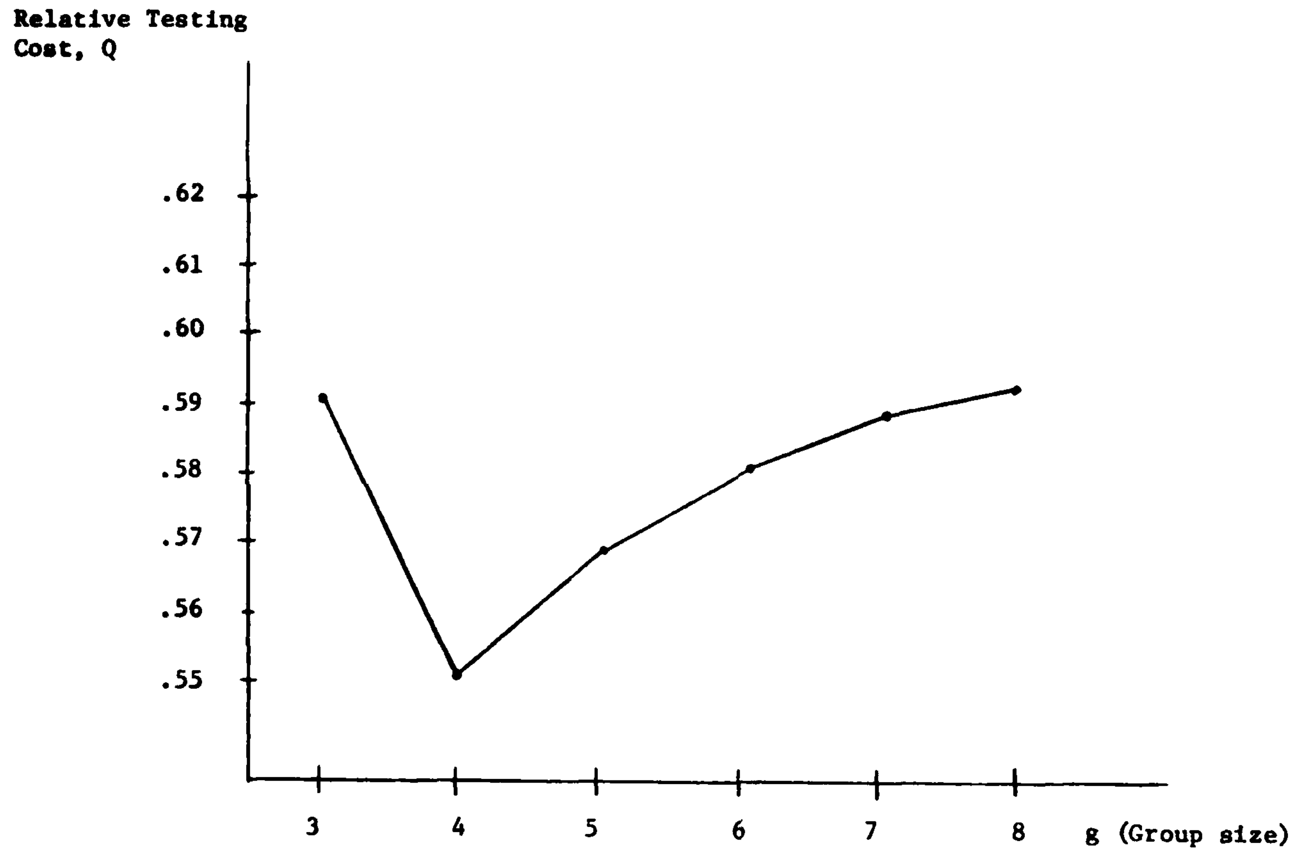

Figure 5: Relative testing cost as a function of group size for the case $K = 120$, $k = 10$ and $\underline{\beta}(3)$.

Appendix B, is:

$$E(R|\underline{\beta}([k/2])) \approx (a + bp^*g)E(R|\underline{\beta}(0)) \approx (a + bp^*g)E_W(R|\underline{\beta}(0)) \ , \tag{6.1}$$

where a = 1.01065 and b = -0.1426. See Figure 6 for an illustration of this approximation.

C. PERFORMANCE

As might be expected, performance is better for smaller p*, i.e., there is more to be gained from the group screening method when the proportion of active factors is smaller. It should be emphasized, however, that for performance (5.4) to be meaningful, both expected loss and relative testing cost must have "acceptable" values. Notwithstanding, a primary result of this paper is that the optimal group size, which maximizes performance under $\underline{\beta}(0)$, is also the optimal group size under $\underline{\beta}(i)$ for $i > 0$. Note that $H_g(\underline{\beta}(0)) = Q^{-1}$. Furthermore, when practical considerations in the number of first stage runs is taken into account (this is not reflected in the data appearing in Appendix B), optimum group size will agree with Watson's optimum group size, setting p = p* in Figure 1, for the case of zero error variance.

In general, performance will decrease as i (in $\underline{\beta}(i)$) increases from 0 to [k/2], although typically the decrease is slight. Figure 7 illustrates this point for the case K = 120 and k = 10. Other cases can be constructed from the data given in Appendix B.

In all, the performance (5.5) of the two-stage group method is fairly insensitive (i.e., no dramatic decrease in performance) to departures from assumption (7). However, it is not entirely clear whether different

Parameter Values	$E(R\mid\underline{\beta}(0))$	$E_w(R\mid\underline{\beta}(0))$	$E(R\mid\underline{\beta}([k/2]))$	$(a + bp^*g)E_w(R\mid\underline{\beta}(0))$
$K = 120$, $k = 16$, $g = 4$ ($p^* = 8/60$)	87.33	86.80	82.31	81.12
$K = 240$, $k = 20$, $g = 7$ ($p^* = 5/60$)	149.0	147.97	137.11	137.23
$K = 120$, $k = 6$, $g = 3$ ($p^* = 3/60$)	63.75	63.61	63.31	62.92
$K = 60$, $k = 2$, $g = 5$ ($p^* = 2/60$)	28.16	27.85	28.02	27.48

Figure 6: Approximating $E(R\mid\underline{\beta}(0))$ and $E(R\mid\underline{\beta}([k/2]))$ with equations (5.7) and (6.1).

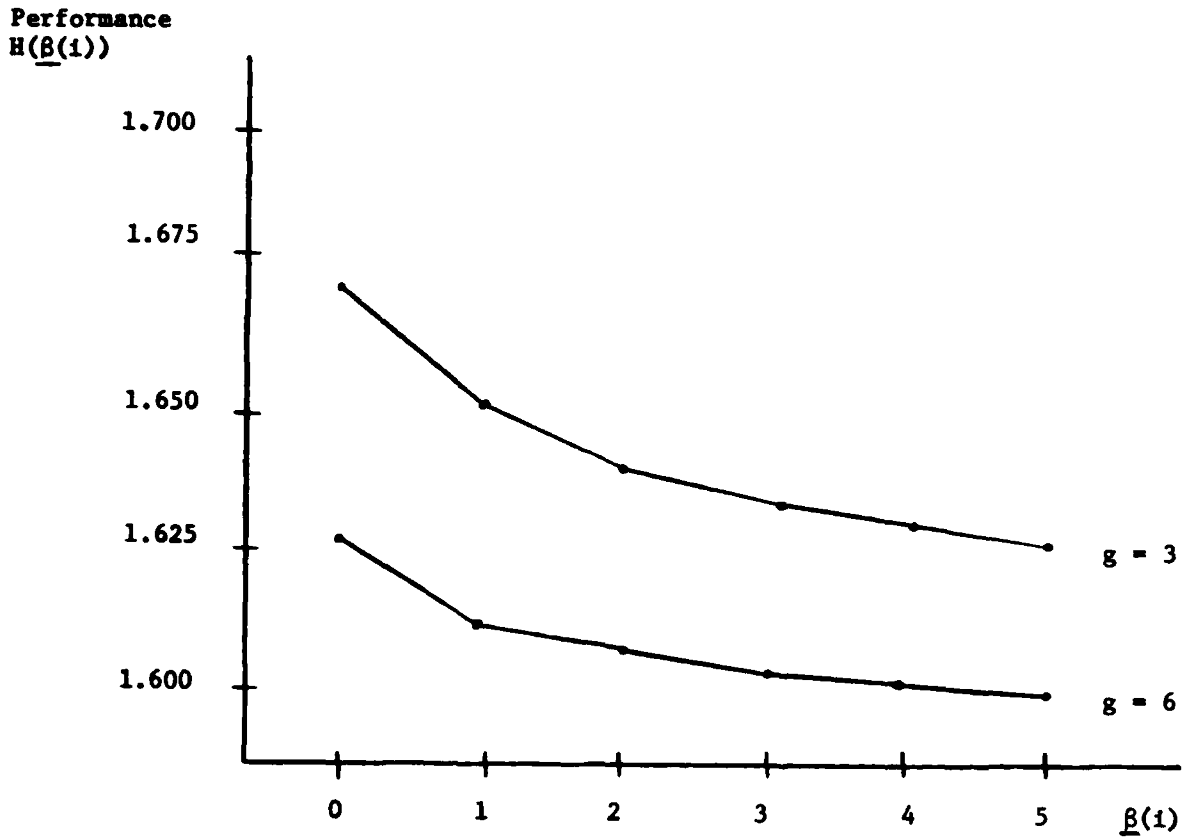

Figure 7: Performance in the case K = 120 and k = 10 for $g = 3$ and $g = 6$.

performance criteria will lead to a common conclusion. This is true, perhaps, of most decision-making problems.

Although measure (5.5) is not altogether comprehensive in evaluating the overall performance of the group screening procedure, no other measure would be either. It must be the decision-maker (i.e., the potential user of group screening) who defines the most appropriate measure for the problem under consideration. In any case, if an overall performance measure is desired, that measure must jointly consider expected loss and relative testing cost.

VII. DISCUSSION

In experimental situations where the factor screening model (4.1) can be assumed, group screening can offer a substantial savings in the required number of experimental runs. However, the possible effects of this economy on factor classification efficiency $[1 - E(L)]$ should be realized especially when the ratio of the number of factors having negative effects to the number of factors having positive effects is near one. In such cases, the risk of factor misclassification may not justify the potential benefits in run economy.

This report explicitly takes into account the impact of possible group-factor effect cancellation on testing cost and on factor classification errors. Accordingly, the results of this report can be used as a practical guide to decisions about the possible use of a group screening strategy, at least in the case $\sigma^2 = 0$.

For experimental situations not included in the twelve cases investigated in this report, testing cost and classification efficiency can be determined from the general formulas presented in Appendix A. In some cases, the data in Appendix B could even be extrapolated to fit the experimental conditions.

Although in the development presented here, $|\beta_i|$, $1 \leq i \leq k$, is assumed equal to Δ, the results of this case can be partially extended to a general $\underline{\beta} = (\beta_1, \beta_2, \ldots, \beta_k)'$ vector of parameters. Indeed, let $\underline{\beta}^*(i)$ denote an arbitrary $\underline{\beta}$ vector having i negative components and (k - i) positive components.

It is easy to see that

$$E(R|\underline{\beta}(i)) \leq E(R|\underline{\beta}^*(i)) \leq E(R|\underline{\beta}(0)) \qquad (7.1)$$

and provided the same loss function (5.2) is used in the $\underline{\beta}^*(i)$ case as in the $\underline{\beta}(i)$ case, it follows that

$$0 = E(L|\beta(0)) \leq E(L|\beta^*(i)) \leq E(L|\underline{\beta}(i)). \qquad (7.2)$$

It should be noted that equation (7.2) may not hold if a loss function other than (5.2) is used in the general $\underline{\beta}$ case. However, equation (7.1) holds no matter what loss function is specified.

Thus, inequalities (7.1) and (7.2) can be used to place limits on loss (if (5.2) is appropriate) and relative testing cost that can occur for an arbitrary $\underline{\beta}$ vector.

VIII. REFERENCES

Hunter, J. S., and Naylor, T. H. (1970). "Experimental Designs for Computer Simulation Experiments," Management Science, Vol. 16, p. 422.

Jacoby, J. E., and Harrison, S. (1962). "Multi-Variable Experimentation and Simulation Models," Naval Research Logistics Quarterly, Vol. 9, p. 121.

Kleijnen, J. P. C. (1975a). "Screening Designs for Poly-Factor Experimentation," Technometrics, Vol. 18, p. 487.

Kleijnen, J. P. C. (1975b). Statistical Techniques in Simulation (Part II), Marcel Dekker, Inc., New York.

Montgomery, D. C. (1979). "Methods for Factor Screening in Computer Simulation Experiments," Technical Report, Georgia Institute of Technology.

Plackett, R. L., and Burman, J. P. (1946). "The Design of Optimum Multifactor Experiments," Biometrika, Vol. 33, p. 305.

Smith, D. E., and Mauro, C. A. (1980). "Factor Screening in Computer Simulation: Considerations in Performance Evaluation," Technical Report No. 113-2, Desmatics, Inc.

Watson, G. S. (1961). "A Study of the Group Screening Method," Technometrics, Vol. 3, p. 371.

APPENDIX A: DERIVATIONS

Suppose $K = Gg$. Let A denote the number of active factors that reach stage two. Let $1 \leq i \leq k-1$, and define

$$c = \begin{cases} 1 \text{ , if factor \#1 carried over to stage two} \\ 0 \text{ , if factor \#1 not carried over to stage two.} \end{cases}$$

Since $i \geq 1$, $\beta_1 = -\Delta$ for the $\underline{\beta}(i)$ case.

Now

$$E(A|\underline{\beta}(i)) = iP[c = 1|\underline{\beta}(i)] + (k - i)P[c = 1|\underline{\beta}(k - i)]$$

$$= i\{1 - P[c = 0|\underline{\beta}(i)]\} + (k - i)\{1 - P[c = 0|\underline{\beta}(k - i)]\}.$$

Thus, it suffices to calculate $P[c = 0|\underline{\beta}(i)]$ for various i.

It is not hard to show that

$$P[c = 0|\underline{\beta}(i)] = W(i)/\binom{K - 1}{g - 1} ,$$

where $W(i) = \sum_{\substack{j=1 \\ j \text{ odd}}}^{\min(k-1,\ g-1)} n_j(i)$

and $n_j(i) = \binom{k - i}{\frac{j + 1}{2}}\binom{i - 1}{\frac{j - 1}{2}}\binom{K - k}{g - 1 - j}$.

Let F denote the number of inactive factors that reach stage two. By a similar argument to the above, it follows that $E(F|\underline{\beta}(i)) =$ $(K - k)(1 - t)$

where $t = Z(i)/\binom{K - 1}{g - 1}$,

$$Z(i) = \sum_{\substack{j=0 \\ j \text{ even}}}^{\min(k-1,\, g-1)} m_j(i) \quad ,$$

and $m_j(i) = \binom{i}{j/2}\binom{k-i}{j/2}\binom{K-k-1}{g-1-j}$.

Regarding the expected total number of runs, $E(R) = E(N + M) = E[B(G) + B(s)] = B(G) + E[B(A + F)]$. Using the approximation $B(x) \cong x + 2.5$, $E(R) \cong B(G) + E(A) + E(F) + 2.5$. Therefore, the performance measure (5.5) may be evaluated as

$$H_g(\underline{\beta}(i)) = \left[\frac{B(K)}{B(G) + E(A|\underline{\beta}(i)) + E(F|\underline{\beta}(i)) + 2.5}\right]\left[\frac{k + E(A|\underline{\beta}(i))}{2k}\right] .$$

APPENDIX B: DATA

The following pages contain the results of the analytical calculations for the twelve specific cases of (K, k) listed in Figure 2. Corresponding to each value of $g(3 \leq g \leq 8)$ and $\underline{\beta}(i)(i = 0, 1, 2, \ldots, [k/2])$ is a block of five numbers. The numbers within a block are arranged as follows:

E(A), the expected number of active factors correctly identified,

E(R), the expected total number of runs,

E(L), the expected loss,

Q, the relative testing cost

and H, the performance measure.

	$\underline{\beta}(0)$	$\underline{\beta}(1)$
	2.000	1.932
	32.40	32.33
g = 3	0.0	0.0169
	0.5062	0.5052
	1.9754	1.9460
	2.000	1.898
	26.30	26.19
g = 4	0.0	0.0254
	0.4109	0.4093
	2.4338	2.3811
	2.000	1.864
	28.16	28.03
g = 5	0.0	0.0339
	0.4400	0.4379
	2.2726	2.2062
	2.000	1.831
	25.99	25.82
g = 6	0.0	0.0424
	0.4061	0.4035
	2.4623	2.3735
	2.000	1.797
	27.79	27.58
g = 7	0.0	0.0508
	0.4342	0.4310
	2.3031	2.2022
	2.000	1.763
	25.55	25.31
g = 8	0.0	0.0593
	0.3992	0.3955
	2.5048	2.3783

K = 60, k = 2, p* = 2/60

	$\underline{\beta}(0)$	$\underline{\beta}(1)$
	3.000	2.867
	35.20	35.00
g = 3	0.0	0.0222
	0.5499	0.5468
	1.8184	1.7881
	3.000	2.804
	29.90	29.50
g = 4	0.0	0.0327
	0.4671	0.4610
	2.1407	2.0982
	3.000	2.743
	32.50	31.86
g = 5	0.0	0.0429
	0.5078	0.4978
	1.9692	1.9228
	3.000	2.684
	31.01	30.06
g = 6	0.0	0.0526
	0.4845	0.4697
	2.0639	2.0169
	3.000	2.628
	33.43	32.12
g = 7	0.0	0.0620
	0.5223	0.5019
	1.9147	1.8688
	3.000	2.575
	31.75	30.05
g = 8	0.0	0.0709
	0.4961	0.4695
	2.0157	1.9788

K = 60, k = 3, p* = 3/60

	$\underline{\beta}(0)$	$\underline{\beta}(1)$	$\underline{\beta}(2)$
	5.000	4.743	4.614
	40.50	40.11	39.92
g = 3	0.0	0.0257	0.0386
	0.6328	0.6268	0.6238
	1.5802	1.5544	1.5413
	5.000	4.635	4.451
	36.54	35.80	35.44
g = 4	0.0	0.0365	0.0549
	0.5709	0.5594	0.5537
	1.7517	1.7222	1.7069
	5.000	4.539	4.307
	40.28	39.13	38.55
g = 5	0.0	0.0461	0.0693
	0.6294	0.6114	0.6024
	1.5888	1.5602	1.5450
	5.000	4.455	4.179
	39.76	38.12	37.30
g = 6	0.0	0.0545	0.0821
	0.6212	0.5957	0.5828
	1.6098	1.5873	1.5750
	5.000	4.382	4.067
	42.97	40.81	39.72
g = 7	0.0	0.0618	0.0933
	0.6715	0.6377	0.6206
	1.4893	1.4713	1.4609
	5.000	4.320	3.966
	41.95	39.23	37.85
g = 8	0.0	0.0680	0.1032
	0.6554	0.6129	0.5913
	1.5257	1.5206	1.5166

K = 60, k = 5, p* = 5/60

	$\underline{\beta}(0)$	$\underline{\beta}(1)$	$\underline{\beta}(2)$	$\underline{\beta}(3)$	$\underline{\beta}(4)$
	8.000	7.575	7.271	7.088	7.027
	47.75	47.11	46.66	46.38	46.29
g = 3	0.0	0.0266	0.0456	0.0570	0.0608
	0.7461	0.7361	0.7290	0.7247	0.7233
	1.3403	1.3223	1.3092	1.3012	1.2985
	8.000	7.429	7.019	6.773	6.690
	45.19	44.05	43.23	42.74	42.57
g = 4	0.0	0.0357	0.0613	0.0767	0.0819
	0.7061	0.6882	0.6755	0.6678	0.6652
	1.4163	1.4011	1.3897	1.3826	1.3802
	8.000	7.320	6.828	6.530	6.430
	49.95	48.25	47.03	46.29	46.04
g = 5	0.0	0.0425	0.0733	0.0919	0.0981
	0.7804	0.7539	0.7348	0.7233	0.7194
	1.2813	1.2701	1.2612	1.2556	1.2536
	8.000	7.243	6.686	6.346	6.231
	50.10	47.83	46.18	45.19	44.85
g = 6	0.0	0.0473	0.0821	0.1034	0.1105
	0.7828	0.7473	0.7216	0.7060	0.7008
	1.2774	1.2748	1.2720	1.2699	1.2692
	8.000	7.192	6.586	6.211	6.083
	53.72	50.89	48.82	47.56	47.13
g = 7	0.0	0.0505	0.0884	0.1118	0.1198
	0.8393	0.7951	0.7628	0.7431	0.7364
	1.1915	1.1942	1.1951	1.1953	1.1952
	8.000	7.165	6.520	6.114	5.976
	52.85	49.51	47.03	45.50	44.99
g = 8	0.0	0.0522	0.0925	0.1179	0.1265
	0.8258	0.7736	0.7348	0.7110	0.7029
	1.2109	1.2251	1.2350	1.2408	1.2427

K = 60, k = 8, $p^* = 8/60$

	$\underline{\beta}(0)$	$\underline{\beta}(1)$	$\underline{\beta}(2)$
	4.000	3.901	3.868
	58.20	58.05	58.00
g = 3	0.0	0.0124	0.0165
	0.4693	0.4681	0.4677
	2.1306	2.1096	2.1026
	4.000	3.854	3.805
	49.90	49.61	49.51
g = 4	0.0	0.0183	0.0244
	0.4024	0.4001	0.3993
	2.4849	2.4539	2.4434
	4.000	3.808	3.745
	49.51	49.03	48.87
g = 5	0.0	0.0239	0.0319
	0.3993	0.3954	0.3941
	2.5046	2.4685	2.4563
	4.000	3.765	3.686
	49.02	48.32	48.08
g = 6	0.0	0.0294	0.0392
	0.3953	0.3896	0.3877
	2.5295	2.4910	2.4779
	4.000	3.723	3.630
	48.44	47.47	47.15
g = 7	0.0	0.0347	0.0463
	0.3907	0.3828	0.3802
	2.5598	2.5216	2.5084
	4.000	3.682	3.576
	47.77	46.50	46.08
g = 8	0.0	0.0397	0.0530
	0.3853	0.3750	0.3716
	2.5957	2.5607	2.5485

$K = 120$, $k = 4$, $p^* = 2/60$

	$\underline{\beta}(0)$	$\underline{\beta}(1)$	$\underline{\beta}(2)$	$\underline{\beta}(3)$
	6.000	5.838	5.740	5.708
	63.75	63.51	63.36	63.31
g = 3	0.0	0.0135	0.0216	0.0244
	0.5141	0.5122	0.5110	0.5106
	1.9450	1.9261	1.9146	1.9108
	6.000	5.765	5.624	5.576
	57.02	56.55	56.27	56.17
g = 4	0.0	0.0196	0.0314	0.0353
	0.4598	0.4561	0.4538	0.4530
	2.1746	2.1497	2.1346	2.1295
	6.000	5.697	5.515	5.454
	58.06	57.31	56.85	56.70
g = 5	0.0	0.0252	0.0404	0.0455
	0.4683	0.4621	0.4585	0.4573
	2.1356	2.1092	2.0930	2.0875
	6.000	5.635	5.415	5.341
	58.89	57.79	57.13	56.91
g = 6	0.0	0.0304	0.0488	0.0549
	0.4749	0.4660	0.4607	0.4590
	2.1058	2.0804	2.0646	2.0592
	6.000	5.577	5.321	5.236
	59.50	58.02	57.13	56.83
g = 7	0.0	0.0353	0.0565	0.0637
	0.4798	0.4679	0.4607	0.4583
	2.0841	2.0620	2.0479	2.0431
	6.000	5.524	5.235	5.139
	59.90	58.00	56.85	56.47
g = 8	0.0	0.0397	0.0637	0.0718
	0.4831	0.4677	0.4585	0.4554
	2.0700	2.0531	2.0421	2.0383

$K = 120$, $k = 6$, $p^* = 3/60$

	$\underline{\beta}(0)$	$\underline{\beta}(1)$	$\underline{\beta}(2)$	$\underline{\beta}(3)$	$\underline{\beta}(4)$	$\underline{\beta}(5)$
	10.000	9.718	9.499	9.342	9.248	9.217
	74.28	73.86	73.53	73.30	73.15	73.11
g = 3	0.0	0.0141	0.0251	0.0329	0.0376	0.0392
	0.5991	0.5956	0.5930	0.5911	0.5900	0.5896
	1.6693	1.6552	1.6441	1.6361	1.6313	1.6297
	10.000	9.606	9.299	9.080	8.948	8.904
	70.16	69.38	68.76	68.32	68.06	67.97
g = 4	0.0	0.0197	0.0351	0.0460	0.0526	0.0548
	0.5658	0.5595	0.5545	0.5510	0.5489	0.5482
	1.7673	1.7521	1.7401	1.7313	1.7260	1.7243
	10.000	9.511	9.129	8.855	8.690	8.636
	73.43	72.21	71.26	70.58	70.17	70.03
g = 5	0.0	0.0245	0.0436	0.0573	0.0655	0.0682
	0.5922	0.5824	0.5747	0.5692	0.5659	0.5648
	1.6886	1.6752	1.6643	1.6564	1.6515	1.6499
	10.000	9.431	8.985	8.664	8.471	8.406
	76.14	74.43	73.10	72.14	71.57	71.37
g = 6	0.0	0.0284	0.0508	0.0668	0.0765	0.0797
	0.6140	0.6002	0.5895	0.5818	0.5771	0.5756
	1.6287	1.6186	1.6103	1.6041	1.6002	1.5989
	10.000	9.365	8.864	8.503	8.285	8.212
	78.31	76.09	74.35	73.10	72.34	72.09
g = 7	0.0	0.0317	0.0568	0.0749	0.0858	0.0894
	0.6315	0.6136	0.5996	0.5895	0.5834	0.5814
	1.5835	1.5780	1.5732	1.5694	1.5670	1.5662
	10.000	9.312	8.765	8.368	8.128	8.047
	79.99	77.24	75.07	73.51	72.57	72.26
g = 8	0.0	0.0344	0.0618	0.0816	0.0936	0.0977
	0.6451	0.6229	0.6054	0.5928	0.5853	0.5827
	1.5502	1.5502	1.5498	1.5491	1.5486	1.5485

K = 120, k = 10, p* = 5/60

	$\underline{\beta}(0)$	$\underline{\beta}(1)$	$\underline{\beta}(2)$	$\underline{\beta}(3)$	$\underline{\beta}(4)$
	16.000	15.556	15.170	14.845	14.578
	88.69	88.02	87.44	86.96	86.56
g = 3	0.0	0.0139	0.0259	0.0361	0.0444
	0.7152	0.7099	0.7052	0.7013	0.6980
	1.3981	1.3892	1.3813	1.3745	1.3689
	16.000	15.413	14.903	14.471	14.116
	87.33	86.16	85.14	84.27	83.57
g = 4	0.0	0.0183	0.0343	0.0478	0.0589
	0.7043	0.6948	0.6866	0.6796	0.6739
	1.4199	1.4129	1.4065	1.4011	1.3965
	16.000	15.312	14.711	14.199	13.778
	92.59	90.87	89.38	88.11	87.06
g = 5	0.0	0.0215	0.0403	0.0563	0.0694
	0.7467	0.7329	0.7208	0.7105	0.7021
	1.3392	1.3352	1.3315	1.3282	1.3253
	16.000	15.245	14.579	14.009	13.537
	96.65	94.39	92.41	90.72	89.33
g = 6	0.0	0.0236	0.0444	0.0622	0.0770
	0.7794	0.7612	0.7452	0.7316	0.7204
	1.2830	1.2828	1.2823	1.2818	1.2812
	16.000	15.205	14.497	13.884	13.373
	99.65	96.86	94.42	92.33	90.60
g = 7	0.0	0.0248	0.0470	0.0661	0.0821
	0.8036	0.7812	0.7614	0.7446	0.7306
	1.2444	1.2483	1.2516	1.2542	1.2563
	16.000	15.187	14.453	13.810	13.269
	101.71	98.46	95.59	93.11	91.05
g = 8	0.0	0.0254	0.0483	0.0684	0.0853
	0.8203	0.7941	0.7709	0.7509	0.7343
	1.2191	1.2274	1.2345	1.2406	1.2456

$K = 120,\ k = 16,\ p^* = 8/60$

	$\underline{\beta}(5)$	$\underline{\beta}(6)$	$\underline{\beta}(7)$	$\underline{\beta}(8)$
	14.371	14.222	14.134	14.104
	86.24	86.02	85.89	85.84
g = 3	0.0509	0.0555	0.0583	0.0593
	0.6955	0.6937	0.6927	0.6923
	1.3646	1.3614	1.3595	1.3589
	13.840	13.643	13.524	13.485
	83.02	82.63	82.39	82.31
g = 4	0.0675	0.0737	0.0774	0.0786
	0.6695	0.6663	0.6644	0.6638
	1.3928	1.3902	1.3886	1.3881
	13.449	13.213	13.071	13.024
	86.25	85.67	85.32	85.20
g = 5	0.0797	0.0871	0.0915	0.0930
	0.6956	0.6909	0.6881	0.6871
	1.3230	1.3213	1.3203	1.3200
	13.166	12.900	12.739	12.685
	88.25	87.47	87.00	86.85
g = 6	0.0886	0.0969	0.1019	0.1036
	0.7117	0.7054	0.7016	0.7004
	1.2807	1.2803	1.2800	1.2799
	12.969	12.678	12.502	12.443
	89.24	88.27	87.68	87.49
g = 7	0.0947	0.1038	0.1093	0.1112
	0.7197	0.7118	0.7071	0.7055
	1.2579	1.2590	1.2596	1.2598
	12.839	12.526	12.337	12.274
	89.43	88.27	87.56	87.33
g = 8	0.0988	0.1085	0.1145	0.1165
	0.7212	0.7118	0.7062	0.7043
	1.2495	1.2523	1.2540	1.2546

K = 120, k = 16, p* = 8/60 (continued)

	$\underline{\beta}(0)$	$\underline{\beta}(1)$	$\underline{\beta}(2)$	$\underline{\beta}(3)$	$\underline{\beta}(4)$
	8.000	7.886	7.804	7.755	7.739
	109.80	109.63	109.51	109.44	109.41
g = 3	0.0	0.0071	0.0122	0.0153	0.0163
	0.4500	0.4493	0.4488	0.4485	0.4484
	2.2222	2.2097	2.2009	2.1955	2.1937
	8.000	7.833	7.714	7.642	7.618
	97.12	96.78	96.55	96.40	96.35
g = 4	0.0	0.0104	0.0179	0.0224	0.0239
	0.3980	0.3967	0.3957	0.3951	0.3949
	2.5124	2.4948	2.4821	2.4745	2.4719
	8.000	7.783	7.628	7.535	7.504
	92.22	91.67	91.29	91.05	90.98
g = 5	0.0	0.0136	0.0233	0.0291	0.0310
	0.3779	0.3757	0.3741	0.3732	0.3728
	2.6460	2.6255	2.6108	2.6019	2.5989
	8.000	7.736	7.547	7.433	7.395
	91.10	90.31	89.74	89.40	89.29
g = 6	0.0	0.0165	0.0283	0.0354	0.0378
	0.3734	0.3701	0.3678	0.3664	0.3659
	2.6783	2.6572	2.6419	2.6326	2.6294
	8.000	7.691	7.470	7.337	7.292
	89.78	88.70	87.93	87.46	87.31
g = 7	0.0	0.0193	0.0331	0.0414	0.0442
	0.3680	0.3635	0.3604	0.3585	0.3578
	2.7177	2.6977	2.6831	2.6741	2.6711
	8.000	7.649	7.397	7.246	7.195
	92.26	90.86	89.85	89.25	89.05
g = 8	0.0	0.0220	0.0377	0.0471	0.0503
	0.3781	0.3724	0.3682	0.3658	0.3649
	2.6447	2.6266	2.6133	2.6051	2.6023

K = 240, k = 8, p* = 2/60

	β(0)	β(1)	β(2)	β(3)	β(4)	β(5)	β(6)
	12.000	11.824	11.679	11.567	11.487	11.439	11.423
	120.87	120.60	120.39	120.22	120.10	120.02	120.00
g = 3	0.0	0.0073	0.0134	0.0180	0.0214	0.0234	0.0240
	0.4954	0.4943	0.4934	0.4927	0.4922	0.4919	0.4918
	2.0188	2.0083	1.9997	1.9931	1.9883	1.9854	1.9844
	12.000	11.747	11.539	11.378	11.263	11.193	11.170
	111.28	110.77	110.36	110.03	109.80	109.67	109.62
g = 4	0.0	0.0106	0.0192	0.0259	0.0307	0.0336	0.0346
	0.4561	0.4540	0.4523	0.4510	0.4500	0.4494	0.4493
	2.1927	2.1795	2.1686	2.1600	2.1539	2.1502	2.1489
	12.000	11.676	11.411	11.205	11.058	10.969	10.940
	109.20	108.40	107.73	107.22	106.85	106.63	106.56
g = 5	0.0	0.0135	0.0245	0.0331	0.0393	0.0430	0.0442
	0.4476	0.4442	0.4415	0.4394	0.4379	0.4370	0.4367
	2.2343	2.2207	2.2093	2.2004	2.1939	2.1900	2.1887
	12.000	11.613	11.295	11.048	10.871	10.764	10.729
	110.67	109.50	108.55	107.81	107.28	106.97	106.86
g = 6	0.0	0.0161	0.0294	0.0397	0.0471	0.0515	0.0530
	0.4536	0.4488	0.4449	0.4419	0.4397	0.4384	0.4379
	2.2048	2.1923	2.1817	2.1734	2.1673	2.1637	2.1625
	12.000	11.555	11.190	10.905	10.701	10.578	10.537
	111.68	110.13	108.85	107.86	107.15	106.72	106.58
g = 7	0.0	0.0185	0.0338	0.0456	0.0541	0.0592	0.0610
	0.4577	0.4513	0.4461	0.4420	0.4391	0.4374	0.4368
	2.1847	2.1746	2.1659	2.1590	2.1540	2.1509	2.1499
	12.000	11.503	11.095	10.775	10.546	10.408	10.362
	116.28	114.29	112.66	111.39	110.48	109.93	109.75
g = 8	0.0	0.0207	0.0377	0.0510	0.0606	0.0663	0.0682
	0.4765	0.4684	0.4617	0.4565	0.4528	0.4505	0.4498
	2.0985	2.0908	2.0841	2.0788	2.0748	2.0724	2.0716

K = 240, k = 12, p* = 3/60

	$\underline{\beta}(0)$	$\underline{\beta}(1)$	$\underline{\beta}(2)$	$\underline{\beta}(3)$	$\underline{\beta}(4)$	$\underline{\beta}(5)$
	20.000	19.706	19.443	19.211	19.010	18.840
	141.85	141.41	141.01	140.67	140.37	140.11
g = 3	0.0	0.0073	0.0139	0.0197	0.0248	0.0290
	0.5814	0.5795	0.5779	0.5765	0.5753	0.5742
	1.7201	1.7128	1.7062	1.7004	1.6953	1.6910
	20.000	19.593	19.228	18.906	18.626	18.390
	137.43	136.62	135.89	135.24	134.69	134.21
g = 4	0.0	0.0102	0.0193	0.0274	0.0343	0.0403
	0.5632	0.5599	0.5569	0.5543	0.5520	0.5501
	1.7754	1.7678	1.7609	1.7548	1.7494	1.7448
	20.000	19.498	19.048	18.650	18.305	18.012
	139.76	138.51	137.38	136.39	135.53	134.80
g = 5	0.0	0.0125	0.0238	0.0337	0.0424	0.0497
	0.5728	0.5676	0.5630	0.5590	0.5554	0.5525
	1.7458	1.7396	1.7338	1.7286	1.7241	1.7201
	20.000	19.421	18.900	18.439	18.038	17.697
	144.93	143.19	141.63	140.26	139.06	138.05
g = 6	0.0	0.0145	0.0275	0.0390	0.0491	0.0576
	0.5940	0.5869	0.5805	0.5748	0.5699	0.5658
	1.6836	1.6793	1.6754	1.6718	1.6686	1.6658
	20.000	19.358	18.780	18.266	17.818	17.437
	149.03	146.78	144.77	142.98	141.43	140.11
g = 7	0.0	0.0160	0.0305	0.0433	0.0545	0.0641
	0.6108	0.6016	0.5933	0.5860	0.5796	0.5742
	1.6373	1.6356	1.6341	1.6325	1.6311	1.6299
	20.000	19.309	18.684	18.127	17.640	17.224
	156.14	153.38	150.89	148.69	146.77	145.14
g = 8	0.0	0.0173	0.0329	0.0468	0.0590	0.0694
	0.6399	0.6286	0.6184	0.6094	0.6015	0.5948
	1.5627	1.5633	1.5638	1.5642	1.5644	1.5645

K = 240, k = 20, p* = 5/60

		$\underline{\beta}(6)$	$\underline{\beta}(7)$	$\underline{\beta}(8)$	$\underline{\beta}(9)$	$\underline{\beta}(10)$
		18.700	18.592	18.515	18.468	18.453
		139.90	139.74	139.62	139.55	139.53
g = 3		0.0325	0.0352	0.0371	0.0383	0.0387
		0.5734	0.5727	0.5722	0.5719	0.5718
		1.6874	1.6847	1.6827	1.6815	1.6811
		18.196	18.046	17.938	17.874	17.852
		133.83	133.53	133.31	133.18	133.14
g = 4		0.0451	0.0489	0.0515	0.0532	0.0537
		0.5485	0.5472	0.5464	0.5458	0.5457
		1.7410	1.7381	1.7359	1.7347	1.7342
		17.772	17.586	17.452	17.372	17.345
		134.20	133.74	133.41	133.21	133.14
g = 5		0.0557	0.0604	0.0637	0.0657	0.0664
		0.5500	0.5481	0.5467	0.5459	0.5457
		1.7169	1.7143	1.7125	1.7114	1.7110
		17.418	17.201	17.045	16.951	16.920
		137.22	136.57	136.11	135.83	135.74
g = 6		0.0645	0.0700	0.0739	0.0762	0.0770
		0.5624	0.5597	0.5578	0.5567	0.5563
		1.6634	1.6616	1.6603	1.6595	1.6592
		17.124	16.880	16.705	16.600	16.565
		139.03	138.19	137.59	137.23	137.11
g = 7		0.0719	0.0780	0.0824	0.0850	0.0859
		0.5698	0.5664	0.5639	0.5624	0.5619
		1.6288	1.6279	1.6273	1.6269	1.6268
		16.882	16.615	16.423	16.308	16.269
		143.80	142.76	142.01	141.56	141.41
g = 8		0.0779	0.0846	0.0894	0.0923	0.0933
		0.5893	0.5851	0.5820	0.5802	0.5796
		1.5645	1.5646	1.5645	1.5645	1.5645

K = 240, k = 20, p* = 5/60 (continued)

	$\underline{\beta}(0)$	$\underline{\beta}(2)$	$\underline{\beta}(4)$	$\underline{\beta}(6)$	$\underline{\beta}(8)$
	32.000	31.122	30.362	29.718	29.192
	170.57	169.25	168.11	167.15	166.36
g = 3	0.0	0.0137	0.0256	0.0357	0.0439
	0.6991	0.6937	0.6890	0.6850	0.6818
	1.4305	1.4218	1.4142	1.4077	1.4023
	32.000	30.849	29.850	29.002	28.307
	171.62	169.32	167.33	165.64	164.25
g = 4	0.0	0.0180	0.0336	0.0468	0.0577
	0.7034	0.6940	0.6858	0.6788	0.6732
	1.4217	1.4151	1.4092	1.4041	1.3999
	32.000	30.659	29.486	28.490	27.669
	177.91	174.56	171.65	169.18	167.15
g = 5	0.0	0.0210	0.0393	0.0548	0.0677
	0.7292	0.7154	0.7035	0.6933	0.6850
	1.3715	1.3685	1.3657	1.3632	1.3610
	32.000	30.535	29.245	28.139	27.224
	185.79	181.40	177.57	174.31	171.62
g = 6	0.0	0.0229	0.0430	0.0603	0.0746
	0.7614	0.7435	0.7277	0.7144	0.7034
	1.3133	1.3143	1.3149	1.3154	1.3156
	32.000	30.463	29.095	27.912	26.927
	191.56	186.20	181.49	177.46	174.13
g = 7	0.0	0.0240	0.0454	0.0639	0.0793
	0.7851	0.7631	0.7436	0.7273	0.7137
	1.2737	1.2789	1.2834	1.2871	1.2901
	32.000	30.432	29.018	27.782	26.743
	199.50	193.26	187.75	183.00	179.06
g = 8	0.0	0.0245	0.0466	0.0659	0.0821
	0.8176	0.7921	0.7694	0.7500	0.7338
	1.2230	1.2316	1.2391	1.2455	1.2508

$K = 240,\ k = 32,\ p^* = 8/60$

	$\underline{\beta}(10)$	$\underline{\beta}(12)$	$\underline{\beta}(14)$	$\underline{\beta}(16)$
	28.782	28.490	28.314	28.256
	165.74	165.31	165.04	164.95
g = 3	0.0503	0.0549	0.0576	0.0585
	0.6793	0.6775	0.6764	0.6760
	1.3981	1.3951	1.3933	1.3927
	27.766	27.379	27.146	27.069
	163.17	162.40	161.94	161.79
g = 4	0.0662	0.0722	0.0758	0.0771
	0.6687	0.6656	0.6637	0.6631
	1.3964	1.3939	1.3924	1.3919
	27.028	26.569	26.292	26.200
	165.57	164.43	163.75	163.53
g = 5	0.0777	0.0849	0.0892	0.0906
	0.6785	0.6739	0.6711	0.6702
	1.3593	1.3580	1.3572	1.3569
	26.506	25.990	25.679	25.575
	169.52	168.02	167.11	166.81
g = 6	0.0858	0.0939	0.0988	0.1004
	0.6948	0.6886	0.6849	0.6837
	1.3158	1.3158	1.3159	1.3159
	26.149	25.588	25.248	25.135
	171.53	169.65	168.53	168.15
g = 7	0.0914	0.1002	0.1055	0.1073
	0.7030	0.6953	0.6907	0.6891
	1.2925	1.2941	1.2951	1.2954
	25.918	25.320	24.957	24.835
	175.96	173.73	172.38	171.93
g = 8	0.0950	0.1044	0.1101	0.1119
	0.7212	0.7120	0.7065	0.7046
	1.2549	1.2579	1.2597	1.2603

K = 240, k = 32, p* = 8/60 (continued)

APPENDIX C: EXPECTED LOSS COMPARISON FOR $p^* = .05$

The following pages present a comparison of expected loss, $E_c(L)$, in the case $p^* = k/K = .05$, for values of $c = 1/12, \ldots, 6/12$. The entries here were obtained from the data in Appendix B. Comparisons of $E_c(L)$ for other values of p^* may also be obtained from the data in Appendix B.

g	K = 60	K = 120	K = 240
3	.006[a]	.007[a]	.007
4	.008[a]	.010[a]	.011
5	.011[a]	.013[a]	.014
6	.013[a]	.015[a]	.016
7	.016[a]	.018[a]	.019
8	.018[a]	.020[a]	.021

$E_c(L)$ for $c = 1/12$ ($p^* = .05$)

g	K = 60	K = 120	K = 240
3	.011[a]	.014	.013
4	.016[a]	.020	.019
5	.022[a]	.025	.025
6	.026[a]	.030	.029
7	.031[a]	.035	.034
8	.036[a]	.040	.038

$E_c(L)$ for $c = 2/12$ ($p^* = .05$)

[a]Entry obtained by linear interpolation.

g	K = 60	K = 120	K = 240
3	.017[a]	.018[a]	.018
4	.025[a]	.026[a]	.026
5	.032[a]	.033[a]	.033
6	.040[a]	.040[a]	.040
7	.047[a]	.046[a]	.046
8	.053[a]	.052[a]	.051

$E_c(L)$ for c = 3/12 (p* = .05)

g	K = 60	K = 120	K = 240
3	.022	.022	.021
4	.033	.032	.031
5	.043	.040	.039
6	.053	.049	.047
7	.062	.057	.054
8	.071	.064	.061

$E_c(L)$ for c = 4/12 (p* = .05)

---

[a] Entry obtained by linear interpolation.

g	K = 60	K = 120	K = 240
3	.022[a]	.023[a]	.023
4	.033[a]	.033[a]	.033
5	.043[a]	.043[a]	.043
6	.053[a]	.052[a]	.052
7	.062[a]	.060[a]	.059
8	.071[a]	.068[a]	.066

$E_c(L)$ for $c = 5/12$ ($p^* = .05$)

g	K = 60	K = 120	K = 240
3	.022[a]	.024	.024
4	.033[a]	.035	.035
5	.043[a]	.046	.044
6	.053[a]	.055	.054
7	.062[a]	.064	.061
8	.071[a]	.072	.068

$E_c(L)$ for $c = 6/12$ ($p^* = .05$)

[a]Entry obtained by linear interpolation.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 113-3	2. GOVT ACCESSION NO. AD-A092	3. RECIPIENT'S CATALOG NUMBER 553
4. TITLE (and Subtitle) AN EXAMINATION OF THE PERFORMANCE OF TWO-STAGE GROUP SCREENING FOR USE IN COMPUTER SIMULATION		5. TYPE OF REPORT & PERIOD COVERED Technical Report
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Carl A. Mauro and Dennis E. Smith		8. CONTRACT OR GRANT NUMBER(s) N00014-79-C-0650
9. PERFORMING ORGANIZATION NAME AND ADDRESS Desmatics, Inc. P. O. Box 618 State College, PA 16801		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR 274-317
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Arlington, VA 22217		12. REPORT DATE October 1980
		13. NUMBER OF PAGES 51
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Distribution of this report is unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Factor Screening
Two-Stage Group Screening
Computer Simulation

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

When only a limited number of simulation runs are available, it is usually impossible to thoroughly investigate all of the factors under consideration. Often, though, it is anticipated that only a small subset of the original factors is important. Accordingly, it is desired to screen the factors in order to help identify those that do exert an appreciable effect on the simulation response. Two-stage group screening is one possible strategy that may be used. However, a basic assumption of this strategy is that the directions

(over)

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

20. ABSTRACT (continued)

of possible effects are known, a priori. This report examines, in the case of zero error variance (i.e., when the simulation response is observed without random error), the performance of two-stage group screening when the assumption of known directions is relaxed.